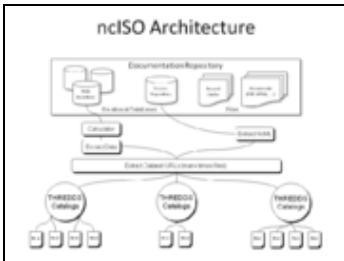


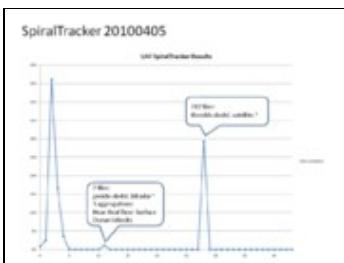
How should data discovery be accomplished?



Discovery Architecture

In the Action Plan Ted's group is building a metadata catalog. Neither the word "discovery" nor "search" is found in our current action plan. What group will offer a user interface services for data discovery? Will this group also offer a service interface for discovery? This diagram shows the current approach being developed for harvesting metadata from the UAF THREDDS catalog.

Discovery Status April 5, 2010



Results

The THREDDS Catalog Crawler illustrated above is now functional. It traverses THREDDS Catalogs and creates three kinds of output:

1. [NetCDF Markup Language](#) exacted directly from the NetCDF file.
2. [ISO 19115-2 XML](#) created from the NcML
3. a SpiralTracker [report](#) for the NcML.

The SpiralTracker report searches the NetCDF file for attributes that are consistent with the [Unidata Data Discovery Conventions](#). SpiralTracker counts compliant attributes and produces a number between 0 and 43.

SpiralTracker was run on 979 files on April 5, 2010 and the distribution of scores is shown in the Figure. There were 287 files with SpiralTracker = 28, which is a really great score. These were [CoastWatch](#) files written by the NOAA CoastWatch, West Coast Node. They are excellent examples of well documented netCDF files. The [ISO](#) and [SpiralTracker Report](#) are also available.

A second group of files had SpiralTracker = 11. These were HF Radar files created by the [National Data Buoy Center](#). The [ISO](#) and [SpiralTracker report](#)

Metadata Sources

NASA Global Change Master Directory

- [NOAA Datasets in the OPeNDAP Portal](#)

UAF_Metadata

- NOAA Datasets in the THREDDS Portal These are listed as being from NSF/UCAR but most are actually NOAA data.